"Release Report"

AMU DEVANAGARI HANDWRITTEN CHARACTER (ADHC) DATASET

at the

Fifth International Conference on Multimedia, Signal Processing, and Communication Technologies (IMPACT-2022)

ABOUT THE DATASET

The ADHC (AMU DEVANAGARI HANDWRITTEN CHARACTER) dataset is a large collection of characters handwritten in Devanagari script. The ADHC Dataset was developed in two versions as a part of two Hackathons on Machine Learning - HACKTELLIGENCE'21 and HACKTELLIGENCE'22, organized by the **Interdisciplinary Centre for Artificial Intelligence**, Aligarh Muslim University, Aligarh, India.

The ADHC datasets were generated by imaging the characters written by many individuals in different writing styles. We scanned these handwritten documents of different writers and cropped each character. Each image in the dataset is unique. The released Dataset is randomly split into Training and Testing set and provided in both CSV and PNG image format.
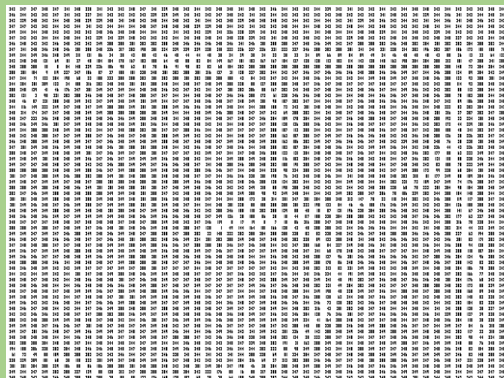
HACKTELLIGENCE'21 DATASET (VERSION 1)

> * This version of the dataset consists of 13,875 images of 10 Swar (Vowels) characters
> * Training set consists of 10,841 images and the testing set consists of remaining 2,944 images
> * around 1500 images for each character
> * Each image has a resolution of 128 by 128
> * 10 different classes as there are 10 Swar characters

Following are the sample images of each character in the dataset.

Moreover, the CSV file is also prepared for every character. One such example for character अ is shown below.

As this version of the dataset consists only Swar (Vovels) characters of the Devanagari script, therefore, it is named as Devanagari Handwritten Swar (DHS) dataset.

HACKTELLIGENCE'22 DATASET (VERSION 2)

- Latest and extended version of DHS dataset
- This version of the dataset consists of 50,325 images of
  - 8 Swar (Vowels) characters
  - 24 Vyanjan (Consonants) characters
- Training set consists of 44,800 images and the testing set consists of remaining 5,525 images
- around 1500 images for each character
- Each image has a resolution of 85 by 85.
- 32 different classes as there are total 32 characters (8 Swar + 24 Vyanjan)

Following are the sample images of each character in the dataset.

.
ACKNOWLEDGEMENT

- We acknowledge the contributions of students of different schools and other individuals who provided their handwritten characters samples for the preparation of these datasets.
- We also acknowledge the contributions made by faculty and students of *Interdisciplinary Centre for Artificial Intelligence, Department of Electronics Engineering, Department of Computer Engineering, Machine Learning Club, and Engineering Design & Implementation Club.*

SOURCE & REFERENCING

The datasets can be cited as follows:

- **Devanagari Handwritten Swar (DHS) dataset,** HACKTELLIGENCE'21 – A College Level HACKATHON on Machine Learning, Interdisciplinary Centre for Artificial Intelligence, A.M.U. Aligarh, December 2021. https://drive.google.com/drive/folders/1nRphsZ7do33waWeGDodiqfKo6xKAm6i1?usp=sharing

  DHS dataset was used in a work on HANDWRITTEN RECOGNITION presented in **IMPACT-2022** and this paper may be cited as follows:

  Mohd Saqib Ansari, Mohammed Wasid, and Syed Atiqur Rahman, "Devanagari Handwritten Character Recognition using Transfer Learning with Deep CNN and SVM", In Proceedings of the *Fifth International Conference on Multimedia, Signal Processing, and Communication Technologies* (IMPACT-2022), IEEE, 2022

- **AMU Devanagari Handwritten Character (ADHC) dataset,** HACKTELLIGENCE'22 – A National Level HACKATHON on Machine Learning, Interdisciplinary Centre for Artificial Intelligence, A.M.U. Aligarh, November 2022. https://drive.google.com/drive/folders/1Pguw7PT6doudvxlX7JGJVZk_O0ti4B3X?usp=share_link